SYSTEMATICS

Molecular Phylogeny of Neotropical *Anopheles* (*Nyssorhynchus*) *albitarsis* Species Complex (Diptera: Culicidae)

RICHARD C. WILKERSON, PETER G. FOSTER,[1] CONG LI, AND MARIA ANICE MUREB SALLUM[2]

Walter Reed Army Institute of Research, Department of Entomology, 503 Robert Grant Avenue, Silver Spring, MD 20910-7500

Ann. Entomol. Soc. Am. 98(6): 918–925 (2005)

ABSTRACT A phylogeny was reconstructed for four species belonging to the Neotropical *Anopheles* (*Nyssorhynchus*) *albitarsis* complex using partial sequences from the mitochondrial cytochrome oxidase I (*COI*) and NADH dehydrogenase 4 (*ND4*) genes and the ribosomal DNA ITS2 and D2 expansion region of the 28S subunit. The basis for initial characterization of each member of the complex was by correlated random amplification of polymorphic DNA-polymerase chain reaction (RAPD-PCR) markers. Analyses were carried out with and without an outgroup (*An.* (*Nys.*) *argyritarsis* Robineau-Desvoidy) by using maximum parsimony, maximum likelihood, and Bayesian methods. A total evidence approach without the outgroup, using separate models for "fast" (*COI* and *ND4* position 3) and "slow" (rDNA ITS2 and D2, and *COI* and *ND4* position 1) partitions, gave the best supported topology, showing close relationships of *An. albitarsis* Lynch-Arribálzaga to *An. albitarsis* B and *An. marajoara* Galvão & Damasceno to *An. deaneorum* Rosa-Freitas. Analyses with the outgroup included showed poorer support, possibly because of a long branch attraction effect caused by a divergent outgroup, which caused one of the *An. marajoara* specimens to cluster with *An. deaneorum* in some analyses. The relationship of the above-mentioned result to a separately proposed hypothesis suggesting a fifth species in the complex is discussed.

KEY WORDS Culicidae, *Anopheles albitarsis* Complex, molecular phylogeny

Anopheles (*Nyssorhynchus*) *marajoara* Galvão & Damasceno, a member of the Albitarsis Complex, was recently recognized as the primary vector of malaria parasites in northeastern Amazonia, Brazil (Conn et al. 2002). Other species in the complex are *An. albitarsis* Lynch-Arribálzaga, *An. deaneorum* Rosa-Freitas, and an unnamed species "B" (Kreutzer et al. 1976; Linthicum 1988; Narang et al. 1993; Rosa-Freitas and Deane 1989; Wilkerson et al. 1995a, b). From studies carried out in Rondônia State, Brazil, there is also evidence to support the importance of *An. deaneorum* as an important malaria vector (Klein et al. 1991a, b). Wilkerson et al. (1995a, b) separated the four largely isomorphic species by using species-specific random amplified polymorphic DNA (RAPD-PCR) markers. Their analysis relied on an empirical assumption of multiple correlated "fixed" markers to hypothesize reproductive isolation, similar to the use of correlated morphological characters for the same ends, but they did not address confounding factors inherent in RAPD markers such as possible nonhomology of comigrating bands or linkage of markers. Lehr et al. (2005), based on complete mitochondrial DNA cytochrome oxidase I (mtDNA *COI*) sequence, question the validity of this approach, suggest a fifth species, and present a *COI* gene tree for all five. Using specimens from the Wilkerson et al. (1995a, b) RAPD-PCR studies, Merritt et al. (2005) analyzed a portion of the *white* gene that contains the *white* fourth intron. They found that the intron was present in *An. marajoara* but not in the other three species (also noted by Krzywinski et al. (2001) for *An. albitarsis*). Phylogenetic analysis of coding sequence in the area of the fourth intron, correlated with intron loss hypotheses, resulted in strong support for a single intron loss event in this species complex and gave a topology different from that found both by Lehr et al. (2005) and from that presented here.

Anopheles subgenus *Nyssorhynchus* includes 33 species (Harbach 2004), including the two most important vectors in the New World tropics: *An. darlingi* Root (Linthicum 1988) and *An. albimanus* Wiedemann (Faran 1980). The subgenus is divided into three sections based on morphological characters: the Argyritarsis Section, which includes the Albitarsis

This material reflects the views of the authors and should not be construed to represent those of the Department of the Army or the Department of Defense.

[1] The Natural History Museum, Department of Zoology, Cromwell Rd., London SW7 5BD, United Kingdom.

[2] Núcleo de Pesquisas Taxonômica e Sistemática em Entomologia Médica, Departamento de Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo, Av. Dr. Arnaldo, 715, 01246-904, São Paulo, SP, Brazil.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
NOV 2005		**00-00-2005 to 00-00-2005**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Molecular Phylogeny of Neotropical Anopheles (Nyssorhynchus) albitarsis Species Complex (Diptera: Culicidae)	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Walter Reed Army Institute of Research,Department of Entomology,Silver Spring,MD,20910**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Table 1. List of species and source of specimens used in this study

Species	Tree reference no	Collection no.	Locality of specimen	Coordinates	GenBank accession no. (AY prefix)			
					ITS2	D2	*ND4*	*COI*
Ingroup								
An. albitarsis	A2	BR511(3)	Paraná, near Guaira	24° 04′ S 54° 15′ W	828322	846323	846340	846307
An. albitarsis	A3	BR501(38)	São Paulo, 6 km SW Registro	24° 36.8′ S 47° 53.1′ W	828322	846324	846341	846308
An. albitarsis	A4	BR500(16)	São Paulo, 6 km SW Registro	24° 36.8′ S 47° 53.1′ W	828322	846325	846342	846309
An. albitarsis	B5	BR017(4)	Bahia, Itaquara	13° 26′ S 39° 56′ W	828324	846326	846343	846310
An. albitarsis	B6	BR/RIO015(1)	Bahia, Itaquara	13° 26′ S 39° 56′ W	828324	846327	846344	846311
An. albitarsis	B7	BR500(45)	São Paulo, 6 km SW Registro	24° 36.8′ S 47° 53.1′ W	828324	846328	846345	846312
An. albitarsis	B8	BR/RIO015(3)	Bahia, Itaquara	13° 26′ S 39° 56′ W	828324	846329	846346	846313
An. marajoara	C9	BR020(8)	Mato Grosso, Peixoto de Azevedo	10° 23′ S 54° 54′ W	828339	846330	846341	846314
An. marajoara	C10	BR026(12)	Amazônas, Manaus	2° 53′ S 60° 15′ W	828339	846331	846348	846315
An. marajoara	C11	BR544(83)	Mato Grosso, Peixoto de Azevedo	10° 23′ S 54° 54′ W	828339	846332	846349	846316
An. marajoara	C12	BR/R001(3)	Pará, Ilha de Marajó	1° 00′ S 49° 30′ W	828328	846333	846350	846317
An. deaneorum	D15	BR/RIO007(11)	Rondônia, Guajará-Mirim	10° 50′ S 65° 20′ W	828332	846334	846351	846318
An. deaneorum	D16	BR700(6)	Rondônia, Ariquemes	9° 56′ S 63° 04′ W	828330	846335	846352	846319
An. deaneorum	D17	BR/RIO007(17).D17	Rondônia, Guajará-Mirim	10° 50′ S 65° 20′ W	828330	846336	846353	846320
Outgroup								
An. argyritarsis	Anargy18	VZ31-100	Venezuela, Merida, Rd to Jaji	7° 37.31′ N 72° 25.92′ W	849553	846337	846354	846321
An. argyritarsis	Anargy19	VZ31-6	Venezuela, Merida, Rd to Jaji	7° 37.31′ N 72° 25.92′ W	849553	846338	846355	846321
An. argyritarsis	Anargy20	BR10-112	Ceará, Ubajara	3° 53.27′ S 40° 54.25′ W	849554	846339	846356	846322

All localities are in Brazil unless otherwise indicated.

Complex and *An. darlingi*; the Albimanus Section, which includes *An. albimanus*; and the little known Myzorhynchella Section (Peyton et al. 1992). Relationships among the 33 included species are not well resolved, but it is known that the Argyritarsis and Albimanus Sections are paraphyletic relative to each other because one putative clade contains both *An. darlingi* and *An. albimanus*, suggesting a possible evolutionary link to vector capacity (Conn 1998, Sallum et al. 2000). Cryptic species are common in *An.* (*Nyssorhynchus*) and in *Anopheles* in general, and most groups that are closely studied yield new taxa (Rosa-Freitas et al. 1998), with *An. darlingi* being an apparent exception (Manguin et al. 1999). Considering the medical importance of this complex and our general lack of knowledge regarding the relationships within subgenus *Nyssorhynchus*, we undertook this study to corroborate results produced by RAPDs and to investigate how the Albitarsis Complex species are related to each other. We report here a molecular phylogenetic analysis of the four species, initially separated by RAPDs, by using two ribosomal DNA (rDNA) sequences, internal transcribed spacer two (ITS2) and the D2 expansion of the 28S subunit (D2), and partial sequence from two mitochondrial genes, NADH dehydrogenase 4 (*ND4*) and *COI*.

Materials and Methods

Source and Identification of Specimens. Morphological characters from Linthicum (1988) were used for identification of *An.* (*Nys.*) *albitarsis* s.l. and *An.* (*Nys.*) *argyritarsis* Robineau-Desvoidy, the outgroup species. The ingroup specimens also were used by Wilkerson et al. (1995a, b) for their studies (Table 1). These were identified to species using species-specific RAPD markers as described by them. For DNA analysis, we used individuals from progeny broods preserved in 100% ethyl alcohol maintained at −70°C. A portion of each brood was retained for morphological study and includes individually reared pinpointed adults with associated pupal and fourth instar exuviae that are deposited in the Smithsonian Institution, National Museum of Natural History (NMNH) and the Faculdade de Saúde Pública, Universidade de São Paulo (FSP-USP). DNA voucher specimens are deposited in NMNH.

Laboratory Methods. DNA was extracted from individual mosquitoes of each species as described in Wilkerson et al. (1993). Portions of the mitochondrial *COI* and the *ND4* genes and the nuclear rDNA, 28S D2 expansion region, and the ITS2 were amplified and sequenced for at least three individuals of each species. The region of each gene that was sequenced, sequences, and positions of the primers used in this study are in Table 2.

PCR reactions were carried out in a total volume of 50 μl by using standard protocols (Palumbi 1996). PCR temperature profiles to obtain the above-mentioned sequence were initial denaturation at 95°C for 3 min followed by 40 cycles of denaturation at 94°C for 40 s; annealing at 56°C for 40 s (ITS2 and *COI*) or 52°C (D2 and *ND4*); and extension at 72°C for 1 min and final extension at 72°C for 10 min.

For sequencing, PCR products were purified using polyethylene glycol (PEG) precipitation (20% PEG 8000 and 2.5 M NaCl). Sequencing reactions were carried out directly on both strands of DNA by using ABI Big Dye chemistry (Applied Biosystems, Foster City, CA), and the sequences were generated with an ABI 377 automated sequencer. The sequences were analyzed and questionable base calls resolved using Sequencher 3.0 (Gene Codes Corp., Ann Arbor, MI). Sequences were initially aligned using ClustalX, ver-

Table 2. Sequences of *COI*, *ND4*, D2, and ITS2 primers used in this study

Designation	Sequence (5'-3')	Position	Reference
Fly5IP(*COI*)	GGATTATTAGGATTTATTGT	842–861[a]	Sallum et al. (2002)
Fly10IP(*COI*)	GCAAATAATGAAATTGTTCT	1373–1392[a]	Sallum et al. (2002)
ND4F	CCAGAACTAATAAAAATCACCAT		This study
ND4R	GCAGGAGTTTTATTAAAGTTAGG		This study
D2F	AGTCGTGTTGCTTGATAGTG	288–307[b]	Sallum et al. (2002)
D2R	CTTGGTCCGTGTTTCAAGAC	821–840[b]	Sallum et al. (2002)
ITS2F	TGTGAACTGCAGGACACATGAA	3' end 5.8S	Cornel et al. (1996)
ITS2R	ATGCTTAAATTTAGGGGGTAGTC	5' end 28S	Cornel et al. (1996)

[a] Nucleotide position relative to *COI* and *COII* sequence of *An. quadrimaculatus* (NC000875), except for Fly10 whose position is relative to *COI* gene in *Drosophila yakuba*.

[b] Nucleotide position relative to 28S sequence of *An. albimanus*.

sion 1.8 (Thompson et al. 1997) and then compared and visually aligned using Se-Al version 2.0a9 (Sequence Alignment Editor, A. Rambaut, University of Oxford) or MacClade (Maddison and Maddison 2000). GenBank accession numbers for all sequences are in Table 1.

Phylogenetic Analyses. Unweighted parsimony analyses were done using PAUP 4.0b10 (Swofford 2004) by using the heuristic search option with TBR branch-swapping with 1000 random-taxon-addition replicates. Parsimony bootstrapping (Felsenstein 1985) used 1000 pseudoreplicates, with 10 random-taxon-addition replicates per pseudoreplicate. Parsimony-uninformative characters and the hypervariable sites were excluded from all the analyses.

For maximum likelihood (ML), we used PAUP 4.0b10 (Swofford 2004). Starting models were chosen with ModelTest (Posada and Crandall 1998) by using the Akaike Information Criterion (AIC). The resulting tree was saved and also used to test site-specific models. Some maximum likelihood determinations and data manipulations were done using p4 (Foster 2004). Maximum likelihood searches started with a neighbor-joining tree, on which we then optimized parameters and fixed the values for those parameters for branch swapping on that tree. Several ML search rounds were carried out until the parameters were fully optimized.

The program p4 (Foster 2004) was used to bootstrap the data, which allowed bootstrapping of partitioned data under a site-specific model. A consensus of the trees from 200 bootstrap replicates was made in PAUP. Branch lengths of that consensus tree were optimized using the search model and the original data, and bootstrap support values from the PAUP tree bipartitions table were placed on the tree using p4.

For Bayesian analysis, we used the program MrBayes (Huelsenbeck and Ronquist 2001). Where possible, the same model used for ML analysis was used in Bayesian analysis, however when that model was not implemented in MrBayes then the next more complex model available in MrBayes was used. When a site-specific model was used, we used site rates, rate matrices, compositions, and among-site rate variation specific to each partition. Markov chain Monte Carlo (MCMC) runs were 500,000 generations long, sampling every 250 generations, for a total of 2,001 samples. Of these, the first 1,001 sample were discarded as burn-in, which is well past the point where the likelihood plot reached a plateau.

To estimate likely placements of the root for the four taxa, we used p4 to compile the sampled MCMC trees with outgroup attachment information preserved to show the distribution of root positions.

Results

In total, 1,846 sites were included in the analysis (Table 3). Because the outgroup was very divergent, we suspected that it might cause us to choose inappropriate models or be responsible for long branch effects. To test this, the analyses were conducted with and without the outgroup. To better fit models to the data, we reasoned that the available genes could be separated into two groupings based on apparent relative substitution rates, "fast" and "slow", with the

Table 3. Constant, variable, and parsimony informative sites in the ingroup only and with the outgroup included

Partition	nSite		nConstSite		nVariableSite		nParsInformativeSite	
	with	without	with	without	with	without	with	without
ITS2*	473	473	449	463	59	10	53	7
D2	559	559	544	553	15	6	9	4
*COI*Pos1	142	142	135	137	7	5	6	4
*COI*Pos2	142	142	141	142	1	0	1	0
*COI*Pos3	143	143	78	103	65	40	51	24
*ND4*Pos1	129	129	121	124	8	5	3	2
*ND4*Pos2	129	129	128	129	1	0	1	0
*ND4*Pos3	129	129	64	80	65	49	55	38

* Hypervariable regions excluded; includes 91 bases in flanking regions.

Table 4. ML boostrap support and posterior probability (PP) for relationships within the *An. albitarsis* complex, when the outgroup was both included and excluded

Phylogenetic group	*COI* + *ND4* pos 3				ITS2+D2+ (*COI* + *ND4* pos 1)				rDNA + mtDNA			
	with		without		with		without		with		without	
	ML	PP	ML	PP	ML	PP	ML	PP	ML	PP	ML	PP
An. albitarsis	0.69	0.93	0.96	1.00	0.53	0.80	0.76	0.65	0.92	0.89	1.00	1.00
An. albitarsis B	0.88	0.93	0.99	1.00	0.77	0.81	0.88	0.96	0.99	0.96	1.00	1.00
An. deaneorum					0.88	0.91	0.81	0.80	0.83	0.95	0.83	0.76
An. marajoara					0.62	0.78	0.91	0.97	0.52	0.64	0.99	1.00
An. marajoara (C9, C11, C12)	0.89	1.00	0.86	1.00					0.58	0.94	0.69	0.93
An. deaneorum (D15, D16)	0.33	0.63	0.48	0.68				0.45	0.73	0.69	0.65	0.67
(*An. marajoara*, *An. deaneorum*)	0.51	0.59	0.99	1.00	0.25	0.52	0.57	0.63	0.66	0.63	1.00	1.00
(*An. albitarsis*, *An. albitarsis* B)	0.51	0.59	0.99	1.00	0.25	0.40	0.76	0.63	0.66		1.00	1.00
(*An. marajoara*, (C10, D17))	0.32	0.37	0.52	0.65								
(*An. marajoara* (C10), *An. deaneorum* (D17)	0.25	0.45	0.63	0.86								

mtDNA position 3 partition (fast) in one group and the remaining, mtDNA position 1 and rDNA (ITS2, D2), in the other group (slow). The data were not subdivided further because of lack of variation. The data groupings as described above as well as the number of parsimony informative characters are given in Table 3.

Phylogenetic Analysis with the Outgroup Excluded. A summary of branch supports for ML and Bayesian analysis for fast partition, slow partition, and combined data are given in Table 4. The best resolution was obtained with combined data (Fig. 1).

Brief mention will first be made here of the separate analyses of fast (mtDNA position 3) and slow (mtDNA position 1 plus rDNA) partitions (Table 4). Using the mtDNA position 3 partition (fast) only, the topologies of the ML and Bayesian analyses were the same, and showed resolution of *An. albitarsis* and *An. albitarsis* B as separate groups, and good resolution of the (*An. deaneorum*, *An. marajoara*) group from the others (designated "A," "B," "D," and "C," respectively in Figs. 1 and 2). However, sequences of *An. marajoara* and *An. deaneorum* were not recovered into two separate nonexclusive clades (de Queiroz 1998) because *An. marajoara* (C10) and *An. deaneorum* (D17) clustered together in a poorly supported clade that was closer to *An. marajoara* than to *An. deaneorum*. Using the slow partition by itself (mtDNA position 1 and rDNA), there was more ambiguity, and the topology for the ML analysis differed from the Bayesian analysis. Support for separate groups was generally poor, with the highest support for the branch separating the entire *An. marajoara* sequence-group from the rest of the tree.

Parsimony analysis of the combined rDNA and mtDNA data sets generated six most passinionious trees (MPTs) (not shown). The strict consensus tree generated from those six MPTs recovered three well-supported groups: *An. albitarsis*, *An. albitarsis* B, and (*An. marajoara* + *An. deaneorum*). The latter also were recovered as separate groups but with less support: *An. deaneorum* with 72% and *An. marajoara* with 91% bootstrap support.

ML and Bayesian analyses were carried out using the mtDNA position 3 in 1 partition, and mtDNA position 1 plus rDNA data in another partition. For the unpartitioned data, ModelTest suggested the TVM + I model; however, the TVM + SS, a site-specific model based on the two partitions, gave a better likelihood, showing an increase of 87.6 log units, so the TVM + SS model was used in all ML analyses. The ML topology, including the nonparametric bootstrap support values, is shown in Fig. 1. For Bayesian analysis, the data were partitioned in the same way. The general time-revers-

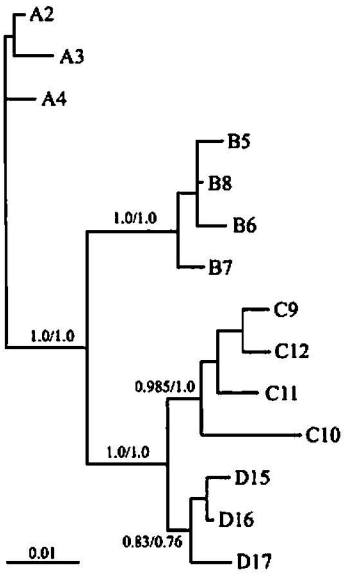

Fig. 1. Results of combined data. The data were placed in two partitions as described in the text, consisting of mtDNA position 3 from COI and ND4 in partition 1, and mtDNA position 1 and ribosomal sequences D2 and ITS2 in the other. The maximum likelihood bootstrap tree and the Bayesian consensus tree had identical topologies, as shown, with support (ML bootstrap/Bayesian posterior probability) shown for the major groups. The ML analysis used the TVM + SS model in PAUP. The Bayesian analysis used a site-specific model, where a GTR + G model was applied to the position 3 mtDNA data partition and GTR + pInvar model to position 1 mtDNA data plus rDNA data partition. A, *An. albitarsis*; B, *An. albitarsis* B; C, *An. marajoara*; D, *An. deaneorum*.

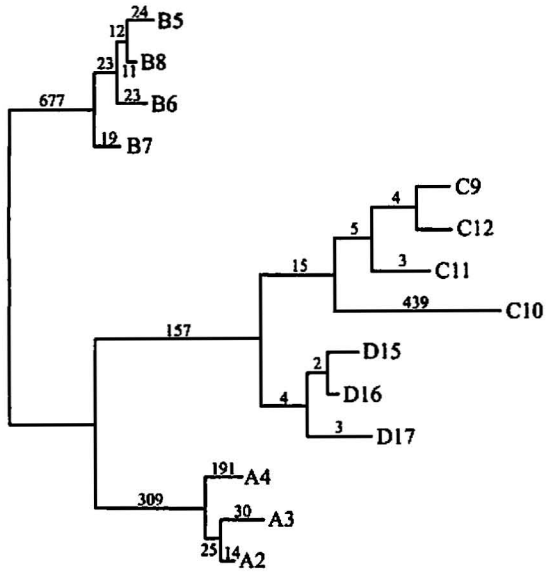

Fig. 2. Outgroup attachment distribution. From postburn-in samples from two MCMC runs using the combined data including the outgroup, a consensus tree was made such that the position of the attachment point of the outgroup was preserved. Of the 2000 samples, most (677) had the outgroup attach as shown, at the base of the *An. albitarsis* B clade. However, during the MCMC the outgroup attachment position varied widely, spending its time on the branches in proportion to the numbers shown. A, *An. albitarsis*; B, *An. albitarsis* B; C, *An. marajoara*; D, *An. deaneorum*.

ible (GTR) rate matrix and composition parameters were unlinked between the two partitions. The applicability of this model was confirmed by ML in p4, which showed an increase of 137 log units by allowing free rates and compositions in the two partitions, at a cost of eight parameters, compared with having the same rate matrix and composition in both partitions. As suggested by ModelTest, gamma-distributed among-site rate variation was applied to the mitochondrial position 3 partition, and a pInvar model was applied to the mitochondrial plus ribosomal partitions. The Bayesian tree has the same topology as the ML bootstrap tree. Bayesian posterior probabilities are shown in Fig. 1. Three groups (*An. albitarsis*, *An. albitarsis* B, and *An. marajoara*) were recovered with high support, and a fourth group, *An. deaneorum*, was recovered with somewhat less support. The support for the split between (*An. albitarsis*, *An. albitarsus* B) and (*An. marajoara*, *An. deaneorum*) was also high.

Phylogenetic Analysis with the Outgroup Included. Here, we note results for the combined data and address the question of where the outgroup attaches to the ingroup (tree topologies not shown). Parsimony analysis (not shown) of the combined rDNA and mtDNA data sets generated 12 MPTs. The strict consensus tree generated from those 12 MPTs recovered one major group consisting of (*An. albitarsis* + *An. albitarsis* B), and two minor groups, one formed by *An. deaneorum* and the other by *An. marajoara* except *An. marajoara* (C10). The relationship between *An. marajoara* and *An. deaneorum* is very poorly supported (<49%), as is the group consisting of *An. marajoara* (<33%). Within the *An. marajoara* clade, *An. marajoara* (C11), *An. marajoara* (C12), and *An. marajoara* (C9) formed a better supported group (72% bootstrap value). Bootstrap support for the clade (*An. albitarsis*, *An. albitarsis* B) is moderate (93%) and for *An. deaneorum* is low (81%).

For the combined data, ModelTest suggested the GTR + I + G model. However, the GTR + SS model, where the data were partitioned into mitochondrial position 3 partition, and mitochondrial position 1 plus ribosomal partitions, had a better likelihood, with an increase of 86 log units, and so the GTR + SS model was used for ML analysis in PAUP. Using ML two major groups were recovered [(*An. deaneorum*, *An. marajoara*), (*An. albitarsis*, *An. albitarsis* B) outgroup]. Support for placement of the outgroup is low (51%), but the groups consisting of *An. albitarsis* and *An. albitarsis* B are moderately and strongly supported (92 and 99%, respectively) (Table 4). For the Bayesian analysis, a site-specific model was used, but a test was made in p4 using maximum likelihood to determine whether individual rate matrices and compositions in the two partitions were better than using a single overall rate matrix and composition. The increase in the log likelihood owing to a separate rate matrix and composition was 205, which is highly significant, and so this strategy was used in the Bayesian analysis. The settings were as described above for combined data without the outgroup. That includes using gamma model for the mitochondrial position 3 partition, and a pInvar model for the ribosomal partition. The support for relationships among the sequences of each ingroup taxon was generally lower when the outgroup was included in both ML and Bayesian analyses than when the outgroup was excluded (Table 4).

To estimate the distribution of likely root positions, the combined postburn-in MCMC samples from two runs were reanalyzed to obtain a consensus tree based on retained root information (Fig. 2). The largest number of input trees (667/2,000) had the outgroup attached on the branch separating the *An. albitarsis* B group from the rest of the tree. However, many other input trees had the outgroup attached on the branches leading to the (*An. marajoara*, *An. deaneorum*) clade, and many other input trees had the outgroup attached on the branch leading to the *An. albitarsis* clade. Also, many trees had the outgroup attached on the branches leading to sequences C10 (*An. marajoara*) and A4 (*An. albitarsis*), which we interpret to be caused by long branch effects and disregard. Few trees had the root attached along the branches leading to either the *An. marajoara* or *An. deaneorum* clades separately, showing little evidence for placement of the root on this part of the tree. However, note that the ML bootstrap tree has this root placement (not shown). Pending further observations we conclude that the rooting is as shown or on the branches leading to the (*An. marajoara*, *An. deaneorum*) clade or to the *An. albitarsis* clade. Stated another way, we find that the largest number of sampled trees had the outgroup attach

between *An. albitarsis* B and [*An. albitarsis* (*An. marajoara, An. deaneorum*)], but with many trees having the outgroup attach at the base of the *An. albitarsis* clade, making [outgroup, (*An. albitarsis*, (*An. albitarsis* B, (*An. marajoara, An. deaneorum*)))] or on the branch separating *An. albitarsis* and *An. albitarsis* B from *An. marajoara* and *An. deaneorum*, making [outgroup ((*An. albitarsis, An. albitarsis* B), (*An. marajoara, An. deaneorum*))]. A topology with *An. albitarsis* B basal to the others was not recovered in any of the other parsimony, ML, or Bayesian consensus trees.

Discussion

In a search for the best evolutionary hypothesis for the Albitarsis Complex, we used partial sequences of two mitochondrial genes (*COI* and *ND4*), and two ribosomal DNA fragments (ITS2 and D2 expansion region of the 28S subunit) and compared maximum parsimony, maximum likelihood, and Bayesian analyses with several combinations of data partitions. Individual genes failed to give well-resolved trees, possibly because of the low number of variable sites. Also, to optimize our results we analyzed different data partition combinations, finally settling on two partitions: mtDNA position 3 alone because it was more variable and presumably faster evolving, and mtDNA position 1 plus rDNA because they were less variable and presumably slower evolving. In addition, because the outgroup was highly divergent, we tested for long branch effects by carrying out all analyses with and without the outgroup.

The strongest support for the evolutionary relationships among the four species tested was retrieved when all four genes were combined and partitioned as described above. Analyses excluding the outgroup, presumably more independent of long branch effects, offer our best hypothesis for the ingroup topology (Fig. 1). In summary, four major evolutive lines were recovered. Two groups, *An. albitarsis*/*An. albitarsis* B and *An. marajoara*/*An. deaneorum*, were usually recovered, but not in all analyses. The latter group includes two species that are important vectors of human *Plasmodium* in localities in the Amazonas region of Brazil (Klein et al. 1991a, b; Conn et al. 2002). This suggests a possible phylogenetic link associated with the ability to transmit human malaria parasites. Similar conclusions have been made for *An. albimanus* and *An. darlingi* (Conn 1998).

When the mtDNA position 3 partition was analyzed alone, *An. deaneorum* specimen D17 and *An. marajoara* specimen C10 clustered together sister to the remaining *An. marajoara* individuals. This result suggests the possibility of incomplete lineage sorting, introgression (Donnelly et al. 2004), or even the existence of an additional taxon.

Our analysis of the distribution of possible roots used sampled trees from MCMC runs, including the outgroup. It showed that *An. albitarsis* and *An. albitarsis* B are in an ambiguous position with relation to the root but that (*An. marajoara, An. deaneorum*) generally formed a clade. Therefore, we think that the root is either as shown in Fig. 2 or as evidenced by the large number of MCMC sampled trees rooting there, either at the base of the *An. albitarsis* B clade, or at the base of the (*An. marajoara, An. deaneorum*) clade. A more definitive answer to this question requires more data.

Lehr et al. (2005) using the entire *COI* gene sequence data of 29 individuals of *An. albitarsis*, *An. albitarsis* B, *An. deaneorum*, and *An. marajoara* recovered results similar to those generated in the current study with the mtDNA position 3 data partition, in that there was nonexclusivity of the *An. deaneorum* clade with respect to *An. marajoara*. They showed the exclusivity of the sequences of *An. albitarsis*, *An. albitarsis* B, and a sister group relationship of these two taxa. They also found four individuals of *An. marajoara* that fell outside the remaining sequences of the Albitarsis Complex (in Bayesian topology), which they suggest represents a fifth species. The individuals that were used to generate these sequences were collected in Roraima State, Boa Vista, Brazil, and Venezuela. Interestingly, Lehr et al. (2005) also recovered a nonexclusive clade consisting of individuals of *An. marajoara* and *An. deaneorum*. A similar grouping was found when we analyzed the mtDNA data partition for the current study (*An. marajoara* specimen C10 clustered with *An. deaneorum* specimen D17). Lack of exclusivity of sequences of *An. marajoara* and *An. deaneorum* are similar to our results and are also suggestive of ancestral introgression or perhaps a recent speciation event that could not be detected by partial sequences of the mitochondrial genes *COI* and *ND4*.

The current study was based on conclusions about species boundaries reached using fixed RAPD markers (Wilkerson et al. 1995a, b). Our results corroborate the RAPD evidence that indicates four putative species: *An. albitarsis*, *An. albitarsis* B, *An. marajoara*, and *An. deaneorum*. No additional taxa were detected. This is not surprising given that the same genetic material was used in both studies. The existence of a fifth species as reported by Lehr et al. (2005) in Boa Vista, Brazil, was not directly tested by us using sequence as described in this study. However, we assumed the Boa Vista specimens to be *An. marajoara* based on comparison to one or two diagnostic RAPD markers (data not shown). The possibility of a fifth species is supported by independent data sets: Kreutzer et al. (1976) (chromosomes), Rosa-Freitas et al. (1990) (isozymes), and possibly Narang et al. (1993) (allozymes, mtDNA restriction fragment length polymorphisms). However, in support of a hypothesis for the existence of *An. marajoara* as a single widespread species is an extensive data set of rDNA ITS2 sequence from throughout its range, and taxon-specific PCR primers based on that sequence (Li and Wilkerson 2005). If in fact there is a fifth member of this complex, the RAPD results should be revisited because assumptions about the wide distribution of *An. marajoara* (Venezuela to southern São Paulo State) were based on the existence of seven RAPD markers that were found in nearly all individuals tested from all parts of its putative range (Wilkerson et al. 1995a, b).

The topology of the gene tree reported by Merritt et al. (2005), who used coding sequence of a portion of the *white* gene containing its fourth intron, varied significantly from that found by us and Lehr et al. (2005). They found good statistical support for a single loss of the fourth intron in the species complex (present in *An. marajoara*, absent in the other species), but weak evidence that *An. marajoara* is basal relative to [*An. albitarsis* B, (*An. albitarsis*, *An. deaneorum*)]. The alternative topology placed *An. marajoara* sister to *An. An. albitarsis* B. There was high support for the sister relationship of *An. albitarsis* and *An. deaneorum*. This is in contrast to our results that give high support for a close relationship between *An. albitarsis* and *An. albitarsis* B and between *An. marajoara* and *An. deaneorum*.

The above-mentioned conflicting results will certainly require additional data to resolve. A recent report by Besansky et al. (2003) addressed the issue of conflicting data sets in the resolution of species boundaries and phylogenetic relationships in the *An. gambiae* complex that may be germane in solving the issue of *An. marajoara*, and resolving the phylogenetic relationships of the Albitarsis Complex. They found evidence supporting both introgression and reproductive isolation as well as different tree topologies, depending on which sequence was sampled. They concluded that adoption of a "total evidence" approach for phylogenetic analysis of closely related species runs a risk of recovering a highly supported wrong answer and suggested that at the level of closely related species, it would be better to do a careful locus-by-locus assessment of sequence divergence rather than just adopt a total evidence approach. They were able to use various genes on all the chromosomes, inside and out of inversions, as well as mitochondrial genes, for their conclusions. The approach of Besansky et al. (2003) provides a model for future research into the phylogenetic relationships of the Albitarsis Complex.

Acknowledgments

We greatly appreciate the support of those who helped collect and rear many of the specimens used in this study, including Terry Klein and Eric Milstrey who were then members of the United States Army Medical Research Unit in Rio de Janeiro. But we especially thank Jose Bento Lima without whose extra efforts many of samples would not have been available to us. We also thank Lee Weigt at the Smithsonian Institution, Laboratory of Analytical Biology, for use of laboratory and sequencing facilities and Jan Conn and Meg Lehr for thoughtful review of a draft of this manuscript. This research was performed under a Memorandum of Understanding between the Walter Reed Army Institute of Research and the Smithsonian Institution, with institutional support provided by both organizations. This work was supported by National Institutes of Helath (AI) R0154139 to J. C.

References Cited

Besansky, N. J., J. Krzywinski, T. Lehmann, F. Simard, M. Kern, O. Mukabayire, D. Fontenille, Y. Taure, and N. F. Sagnon. 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. Proc. Nat. Acad. Sci. U.S.A. 100: 10818–10823.

Conn, J. E. 1998. Systematics and population level analysis of *Anopheles darlingi*. Mem. Inst. Oswaldo Cruz 93: 647–650.

Conn, J. E., R. C. Wilkerson, M.N.O. Segura, T. L. de-Souza-Raimundo, C. D. Schlichting, R. A. Wirtz, and M. M. Povoa. 2002. Emergence of a new Neotropical malaria vector facilitated by human migration and changes in land use. Am. J. Trop. Med. Hyg. 66: 18–22.

Cornel, A. J., C. H. Porter, and F. H. Collins. 1996. Polymerase chain reaction species diagnostic assay for *Anopheles quadrimaculatus* cryptic species (Diptera: Culicidae) based on ribosomal DNA ITS2 sequences. J. Med. Entomol. 33: 109–116.

de Queiroz, K. 1998. Endless forms: species and speciation, pp. 57–75. *In* D. J. Howard and S. H. Berlocher [eds.], The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. Oxford University Press, Oxford, England.

Donnelly, M. J., J. Pinto, R. Girod, N. J. Besansky, and T. Lehmann. 2004. Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. Heredity 92: 61–68.

Faran, M. E. 1980. Mosquito studies (Diptera: Culicidae) XXXIV. A revision of the Albimanus section of the subgenus *Nyssorhynchus* of *Anopheles*. Cont. Am. Entomol. Inst. 15: 1–215.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

Foster, P. G. 2004. Modeling compositional heterogeneity. Syst. Biol. 53: 485–495.

Harbach, R. E. 2004. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. Bull. Entomol. Res. 94: 537–553.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.

Klein, T. A., J.B.P. Lima, and M. S. Tada. 1991a. Comparative susceptibility of anopheline mosquitoes to *Plasmodium falciparum* in Rondonia Brazil. Am. J. Trop. Med. Hyg. 44: 598–603.

Klein, T. A., J.B.P. Lima, M. S. Tada, and R. Miller. 1991b. Comparative susceptibility of Anopheline mosquitoes in Rondonia Brazil to infection by *Plasmodium vivax*. Am. J. Trop. Med. Hyg. 45: 463–470.

Kreutzer, R. D., J. B. Kitzmiller, and M. G. Rabbani. 1976. Cytogenetically distinguishable sympatric and allopatric populations of the mosquito *Anopheles albitarsis*. Acta Amazonica 6: 473–481.

Krzywinski, J., R. C. Wilkerson, and N. J. Besansky. 2001. Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence. Syst. Biol. 50: 540–556.

Lehr, M. A., C. W. Kilpatrick, R. C. Wilkerson, and J. E. Conn. 2005. Cryptic species in the *Anopheles* (*Nyssorhynchus*) *albitarsis* (Diptera: Culicidae) complex: incongruence between random amplified polymorphic DNA-polymerase chain reaction identification and analysis of mitochondrial DNA *COI* gene sequences. Ann. Entomol. Soc. Am. 98: 894–903.

Li, C., and R. C. Wilkerson. 2005. Identification of *Anopheles* (*Nyssorhynchus*) *albitarsis* complex species (Diptera:

Culicidae) using rDNA ITS2-based PCR primers. Mem. Inst. Oswaldo Cruz 100: 495–500.

Linthicum, K. J. 1988. A revision of the Argyritarsis Section of the subgenus *Nyssorhynchus* of *Anopheles* (Diptera Culicidae). Mosq. Syst. 20: 98–271.

Maddison, D. R., and M. D. Maddison. (2000). MacClade 4, analysis of phylogeny and character evolution. Sinauer, Sunderland, MA.

Manguin, S., R. C. Wilkerson, J. E. Conn, P. Y. Rubio, B.J.A. Danoff, and D. R. Roberts. 1999. Population structure of the primary malaria vector in South America, *Anopheles darlingi*, using isozyme, random amplified polymorphic DNA, internal transcribed spacer 2, and morphologic markers. Am. J. Trop. Med. Hyg. 60: 364–376.

Merritt, T.J.S., C. R. Young, R. G. Vogt, R. C. Wilkerson, and J. M. Quattro. 2005. Intron retention identifies a malaria vector within the *Anopheles* (*Nyssorhynchus*) *albitarsis* complex (Diptera: Culicidae). Mol. Phylogenet. Evol. 35: 719–724.

Narang, S. K., T. A. Klein, O. P. Perera, and A. T. Tang. 1993. Genetic evidence for the existence of cryptic species in the *Anopheles albitarsis* complex in Brazil: allozymes and mitochondrial DNA restriction fragment length polymorphisms. Biochem. Gen. 31: 97–112.

Palumbi, S. R. 1996. Nucleic acids II: the polymerase chain reaction, pp. 205–247. *In* D. M. Hillis, C. Moritz, and B. K. Mable [eds.], Molecular systematics. Sinauer, Sunderland, MA.

Peyton, E. L., R. C. Wilkerson, and R. E. Harbach. 1992. Comparative analysis of the subgenera *Kerteszia* and *Nyssorhynchus* of *Anopheles* (Diptera Culicidae). Mosq. Syst. 24: 51–69.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817–818.

Rosa-Freitas, M. G., and L. M. Deane. 1989. The neotype of *Anopheles albitarsis* (Diptera: Culicidae). Mem. Inst. Oswaldo Cruz 84: 289–302.

Rosa-Freitas, M. G., L. M. Deane, and H. Momen. 1990. A morphological isoenzymatic and behavioral study of ten populations of *Anopheles albitarsis* Lynch-Arribalzaga 1878 (Diptera: Culicidae) including from the type locality, Baradero, Argentina. Mem. Inst. Oswaldo Cruz 85: 275–290.

Rosa-Freitas, M. G., R. Lourenço de Oliveira, C. J. de Carvalho-Pinto, C. Flores-Mendoza, T. Fernandes, and D. M. Silva. 1998. Anopheline species complexes in Brazil. Current knowledge of those related to malaria transmission. Mem. Inst. Oswaldo Cruz 93: 651–655.

Sallum, M.A.M., T. R. Schultz, and R. C. Wilkerson. 2000. Phylogeny of Anophelinae (Diptera Culicidae) based on morphological characters. Ann. Entomol. Soc. Am. 93: 745–775.

Sallum, M.A.M., T. R. Schultz, P. G. Foster, K. Aronstein, R. A. Wirtz, and R. C. Wilkerson. 2002. Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. Syst. Entomol. 27: 361–382.

Swofford, D. L. 2004. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24: 4876–4882.

Wilkerson, R. C., T. J. Parsons, D. G. Albright, T. A. Klein, and M. J. Braun. 1993. Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: *Anopheles*). Insect Mol. Biol. 1: 205–211.

Wilkerson, R. C., T. V. Gaffigan, and J. B. Lima. 1995a. Identification of species related to *Anopheles* (*Nyssorhynchus*) *albitarsis* by random amplified polymorphic DNA-Polymerase chain reaction (Diptera: Culicidae). Mem. Inst. Oswaldo Cruz 90: 721–732.

Wilkerson, R. C., T. J. Parsons, T. A. Klein, T. V. Gaffigan, E. Bergo, and J. Consolim. 1995b. Diagnosis by random amplified polymorphic DNA polymerase chain reaction of four cryptic species related to *Anopheles* (*Nyssorhynchus*) *albitarsis* (Diptera: Culicidae) from Paraguay, Argentina, and Brazil. J. Med. Entomol. 32: 697–704.

Received 15 February 2005; accepted 28 July 2005.